



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

**POZITIVNÍ A NEGATIVNÍ SELEKCE
MITOCHONDRIÁLNÍHO GENOMU**

POZITIVE AND NEGATIVE SELECTION OF MITOCHONDRIAL GENOME

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Matěj Svoboda

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Denisa Maděránková, Ph.D.

BRNO 2017

Bakalářská práce

bakalářský studijní obor **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

Student: Matěj Svoboda

ID: 164999

Ročník: 3

Akademický rok: 2016/17

NÁZEV TÉMATU:

Pozitivní a negativní selekce mitochondriálního genomu

POKYNY PRO VYPRACOVÁNÍ:

1) Vypracujte literární rešerši na téma negativní a pozitivní selekce mitochondriálního genomu a metody určování selekce. 2) V programovém prostředí R vytvořte funkci pro hromadnou extrakci zájmových oblastí z celogenomových sekvenčních záznamů. 3) Dále vytvořte funkci pro výpočet parametrů nutných k určení výskytu negativní/pozitivní selekce. 4) Pomocí vytvořených funkcí vyhodnoťte selekci mitochondriálních genomů rozsáhlého souboru bezobratlých i obratlovců. 5) Výsledky porovnejte s výsledky nástroje PAML.

DOPORUČENÁ LITERATURA:

[1] JAMES, J.E.; PIGANEAU, G.; EYRE-WALKER, A. The rate of adaptive evolution in animal mitochondria. *Molecular Ecology*. 2016, 25, 67-78.

[2] MEIKLEJOHN, C.D.; MONTTOOTH, K.L.; RAND, D.M. Positive and negative selection on the mitochondrial genome. *TRENDS in Genetics*. 2007, 23(6), 259-263.

Termín zadání: 6.2.2017

Termín odevzdání: 2.6.2017

Vedoucí práce: Ing. Denisa Maděránková, Ph.D.

Konzultant:

prof. Ing. Ivo Provazník, Ph.D.
předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

Abstrakt

Bakalářská práce se zabývá problematikou pozitivní a negativní selekce mitochondriálního genomu. Práce je rozdělena na dvě části. První část se věnuje teorii a vysvětlení základních pojmů, konkrétně popisuje mitochondriální genom, nukleotidové mutace, pozitivní a negativní selekci a v neposlední řadě evoluční modely. Druhá část je zaměřena na praktické zpracování mitochondriální DNA a vytvoření funkcí v programovacím prostředí R. Důraz je kladen především na zarovnání sekvencí genů a hledání substitucí. Získané výsledky jsou následně porovnávány s výstupy z programů PAML a KaKs Calculator.

Klíčová slova

DNA, evoluční modely, genom, mitochondrie, negativní selekce, PAML, pozitivní selekce, R Studio

Abstract

The bachelor thesis engages with problematics of positive and negative selection of mitochondrial genome. Thesis is divided into two parts. First part grapples with theory and explanation of fundamental definitions, in particular understates mitochondrial genome, nucleotide mutation, positive and negative selection and furthermore evolution models. Second part focuses on processing mitochondrial DNA practically and on establishing functions in programming environment R. The emphasis is foremost on gene sequencing alignment and exploring substitutions. Consequently, obtained outcomes are contrasted with PAML and KaKs Calculator programme outcomes.

Keywords

DNA, evolution model, genome, mitochondria, negative selection, PAML, positive selection, R Studio

SVOBODA, M. Pozitivní a negativní selekce mitochondriálního genomu. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2017. 27s.
Vedoucí práce: Ing. Denisa Maděránková, Ph.D.

Prohlášení

Prohlašuji, že svou závěrečnou práci na téma *Pozitivní a negativní selekce mitochondriálního genomu* jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne **2. června 2017**

.....
podpis autora

Poděkování

Děkuji vedoucímu bakalářské práce Ing. Denise Maděránkové, Ph.D. za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé bakalářské práce.

V Brně dne **2. června 2017**

.....
podpis autora

Obsah

Seznam obrázků	iii
Seznam tabulek	iv
1. Úvod.....	1
2. Mitochondrie.....	2
2.1. Mitochondriální genom.....	3
3. Mutace	5
3.1. Genová mutace.....	6
3.2. Bodové mutace.....	7
3.3. Synonymní mutace.....	7
3.4. Nesynonymní mutace.....	7
3.5. Nesmyslné mutace	8
3.6. Negativní selekce	8
3.7. Pozitivní selekce.....	9
3.8. Chromozomální mutace	9
3.9. Genomová mutace.....	9
4. Zarovnání sekvencí	10
4.1. Zarovnání bez mezer	10
4.2. Zarovnání s mezerami	10
4.3. Vícenásobné zarovnání	10
5. KaKs kalkulátor	11
6. Evoluční modely	12
6.1. Základní evoluční modely.....	12
6.2. Proporcionální vzdálenost sekvencí.....	12
6.3. Neutrální model.....	12
6.4. Jukes-Cantor model.....	13
6.5. Kimura model.....	13
6.6. Felsenstein model.....	14
6.7. Maximum likelihood model.....	14
7. Řešení problematiky v programu R.....	15
8. Výsledky	21
9. Závěr	27
Citovaná literatura.....	29

Seznam zkratek	II
Přílohy	IV
Příloha 1	IV
Příloha 2	V
Příloha 3	VI

Seznam obrázků

Obr. 1: Schéma mitochondrie, převzato z [15]	2
Obr. 2: Cyklická mitochondriální DNA, převzato z [7]	4
Obr. 3: Bodové mutace, [12]	8
Obr. 4: Vícenásobné zarovnání pomocí internetového nástroje ClustalO.....	11
Obr. 5: Kimurův dvouparametrový model	13
Obr. 6: Blokové schéma vytvořených funkcí	16
Obr. 7: Vícenásobné zarovnání aminokyselin	17
Obr. 8: Vývojový diagram funkce NSites	19
Obr. 9: Míra substituce nukleotidů pro gen ND2 organismů <i>Glis glis</i> a <i>Artibeus jamaicensis</i>	24
Obr. 10: Míra substituce nukleotidů pro gen ND2 organismů <i>Glis glis</i> a <i>Cavia porcellus</i>	24

Seznam tabulek

Tab. 1: Odlišnosti genetického kódu a mitochondriálního genomu obratlovců	5
Tab. 2: Míra substitucí nukleotidů, K80	14
Tab. 3: Míry substitucí nukleotidů, F81	14
Tab. 4: Příklad výpočtu funkce NSites pro AK Methionin (Met, ATG)	18
Tab. 5: Výsledné hodnoty funkce KaKs kalkulátor pro gen ND1	22
Tab. 6: Hodnoty NaN.....	23
Tab. 7: Hodnoty analyzovaných sekvencí v programu PAML	25
Tab. 8: Výsledné hodnoty pro Glis Glis, gen ATP8.....	IV
Tab. 9: Výsledné hodnoty pro Glis Glis, gen ND2.....	V
Tab. 10: Pozitivní selekce	VI

1. Úvod

Mitochondrie jsou orgány endosymbiotického původu. Důkazem je přítomnost vlastní dvojité fosfolipidové membrány, a především vlastní mitochondriální DNA podobající se DNA bakteriím. Vlastní DNA zaručuje mitochondriím částečnou autonomii. Dědičnost mtDNA je převážně matroklinní, mitochondriální genom získává potomstvo výhradně od matky. DNA mitochondrií je využívána pro analýzu evoluční příbuznosti, protože nepodléhá rekombinacím DNA při pohlavním rozmnožování.

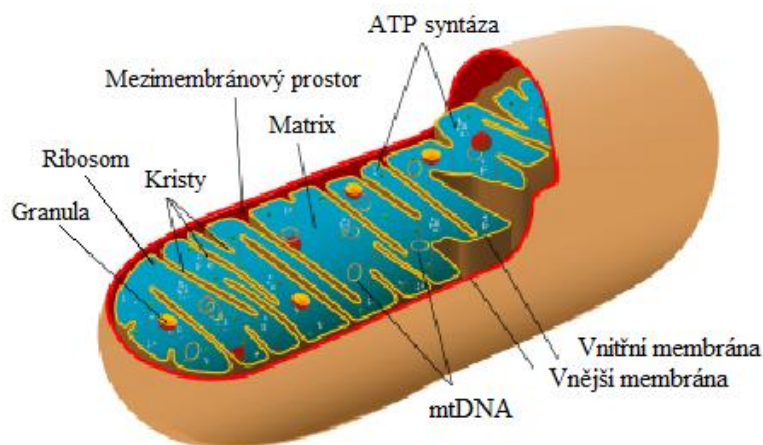
Mutace jsou dědičné změny genotypu organismu. Dochází k nim zcela náhodně, většina je opravena reparačními systémy buňky. Působením mutagenů se šance vzniku mutace zvyšuje. Důsledkem mutací je variabilita genetické informace jedinců příslušících stejnému druhu, jinými slovy, mutace přispívají ke genetické diverzitě a jsou hlavní příčinou evoluce. Bodové mutace lze rozdělit na synonymní, nesynonymní a nesmyslné. Produktem synonymní substituce je protein shodný s původním proteinem, jeho funkce se nemění. Evolučním motorem jsou převážně mutace nesynonymní, měnící funkci výsledného proteinu. Změny funkce mohou být latentní či prospěšné. Distribuce genů v populaci neboli tzv. selekce, může být pozitivní či negativní. Pozitivní selekcí se do populace dostávají prospěšné geny, naopak negativní selekce napomáhá vymizení škodlivých genů.

Evoluční modely se používají v genetice pro definování evoluční vzdálenosti nukleotidových či proteinových sekvencí. Komplexnost výpočtu vzniklých substitucí se liší model od modelu. Nejjednodušším výpočtem evoluční vzdálenosti je proporcionální vzdálenost sekvencí. Nejjednodušším substitučním modelem je model Jukes-Cantor. K složitějším substitučním modelům se řadí Kimurův dvouparametrický model, Felsensteinův model nebo Maximum likelihood. Evoluční modely jsou zahrnuty i v programech jako je například KaKs calculator, nebo PAML, který je využit pro porovnání výsledků v této bakalářské práci.

Bakalářská práce se věnuje analýze sekvencí mitochondriální DNA získané z databáze NCBI a hledá přítomnosti synonymních a nesynonymních mutací a hodnoty jejich poměrů. Výsledky jsou získávány také evolučními modely. Výpočty jsou realizovány vytvořenými funkcemi v programovém prostředí R. Cílem bakalářské práce je zjistit přítomnost pozitivní nebo negativní selekce. Pro porovnání je využito výstupů z programu PAML.

2. Mitochondrie

Mitochondrie je membránová buněčná organela, má kulovitý vláknitý nebo



Obr. 1: Schéma mitochondrie, převzato z [15]

válcovitý tvar a svými rozměry připomíná bakterie. Nalézá se v eukaryotických buňkách v cytoplazmě. Její hlavní funkcí je produkce adenosintrifosfátu (ATP), který slouží jako zdroj energie pro organismus. ATP je tvořeno procesem oxidativní fosforylace. Tímto se mitochondrie stává nepostradatelnou, protože tvoří většinu energie pro buňky. Zastoupení mitochondrií je v různých tkáních odlišné. Nejvíce je jich obsaženo v buňkách s intenzivním metabolismem, jako jsou svaly, neurony, anebo v jaterní tkáni. Mitochondrie se skládá ze dvou membrán. Vnější membrána je hladká, zatímco vnitřní membrána vytváří záhyby, kristy, do mitochondrie. Prostor v mitochondrii je membránami rozdělen na dva oddíly, vnější a vnitřní (matrix). Enzymy spjaté s respirací jsou zabudovány do vnitřní mitochondriální membrány. Mitochondrie se reprodukují příčným dělením, podobně jako se rozmnožují bakterie a také respirační procesy jsou podobné procesům volně žijícím aerobním bakteriím.

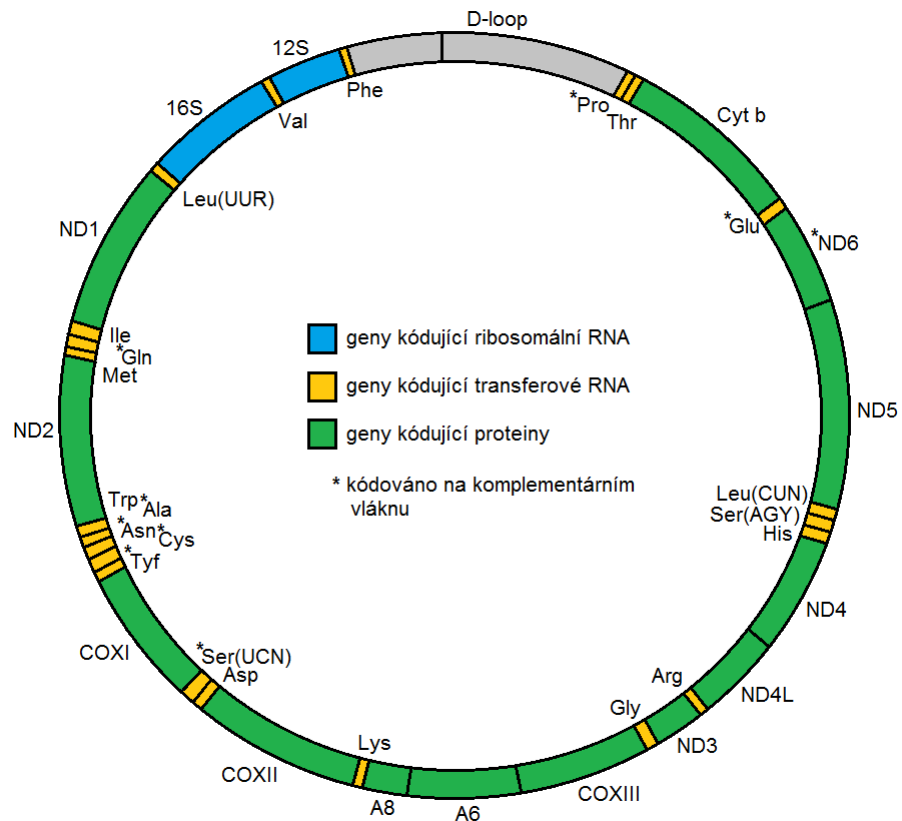
Mezi obratlovci, rostlinami, houbami a dalšími se mitochondriální genom liší například počtem kódovaných genů, délkou řetězce mitochondriální DNA (mtDNA) nebo také expresí genů. Mitochondriální DNA *Reclinomonas americana* obsahuje 97 různých genů kódujících proteiny oproti tomu mtDNA *Plasmodium falciparum* má tyto geny pouze 3. U zvířat a člověka je mitochondrie řízena svými vlastními geny a geny z jádra buňky. I přes tyto odlišnosti zůstává syntéza proteinů nezměněná, [1, 14].

Geny zakódované v mtDNA se podílí na translaci, oxidativní fosforylaci, při transportu elektronů a na mimojederové dědičnosti.

2.1. Mitochondriální genom

Většina genetické informace buňky se nalézá v jádře, a proto ji nazýváme jaderná DNA. V roce 1949 byla objeveno, že mitochondrie mají svou vlastní DNA (mtDNA), jedná se tedy o jedinou organelu, kromě jádra, která obsahuje vlastní genetickou informaci v savcích buňkách. Mitochondriální DNA je cirkulární dvou řetězcová molekula. Stejně jako jiné DNA skládá se z nukleotidů adeninu (A), guaninu (G), cytosinu (C) a thyminu (T). V ribonukleové kyselině (RNA) je thymin nahrazen uracilem (U). V jedné organelle je více kopií mtDNA. U většiny zvířat a u člověka má genom mitochondrie okolo 16 kb. Výrazně větší genom mají zástupci kvasinek (60 kb), anebo rostlin (200 kb a více). Takto dlouhé sekvence z velké části obsahují nekódující úseky tzv. introny. Všechny ribosomální RNA (rRNA) a většina transferových RNA, potřebných k translaci mitochondriálního genomu, jsou zakódovány v mitochondriálním genomu. Zbytek proteinů obsažených v mitochondrii je zakódován v genech jádra buňky. Tyto informace se do jádra pravděpodobně přenesly z předka současných mitochondrií, [2].

Většina proteinů působících v mitochondriích je kódována jadernou DNA, translace probíhá v cytoplazmě a následně jsou proteiny transportovány do mitochondrie. V samotné mtDNA je zakódováno dalších 13 proteinových struktur, syntéza probíhá na speciálních mitochondriálních ribozomech. Tyto proteiny jsou důležité pro proces oxidativní fosforylace, naváží se na komplexy oxidativní fosforylace a dostanou se do vnitřní membrány mitochondrie. Z toho vyplývá, že mitochondrie jsou řízeny jak jádrem, tak i samy sebou. Jinak řečeno podléhají duální genetické kontrole, to z nich dělá semiautonomní organely, [14].



Obr. 2: Cyklická mitochondriální DNA, převzato z [7]

V mtDNA je malé množství kodonů, které kódují odlišné aminokyseliny (AK) v mitochondriálním genetickém kódu a jiné AK v univerzálním genetickém kódu. V jsou tyto rozdíly znázorněny. Například start kodon mitochondrií není pouze triplet AUG (Methionin), ale je pět možných alternativ. V tabulce jsou znázorněny zelenou barvou. Stejně tak tomu je i u stop kodonů. Jsou čtyři a jsou v tabulce znázorněny červenou barvou. Mnoho molekul transferové RNA (tRNA) (u prokaryot i eukaryot) je schopno se spárovat s více než jedním kodonem mediátorové RNA (mRNA), protože mezi třetí pozicí kodonu a antikodonem tRNA občas dojde k chybnému párování. K translaci univerzálního genetického kódu je zapotřebí minimálně 30 různých tRNA. Mitochondrie savců má ve své DNA zakódováno pouze 22 různých typů tRNA, které následně vstupují do translace. Z toho vyplývá že, jedna molekula mitochondriální tRNA musí být schopna přechíst všechny kodony, které kódují stejnou amino kyselinu nebo top kodon, který způsobí konec translace (např. GGU, GGC, GGA a GGG kódují stejnou AK (glycin) nebo UAA, UAG a UGA kódují stop kodon). V některých případech není jasné, zda k párování stačí dva nukleotidové páry, nebo je uridin na první pozici antikodonu schopný párovat se všemi čtyřmi nukleotidy, které jsou na třetí pozici kodonu. Díky tomuto jevu je jedna molekula tRNA schopna rozpoznat 4 různé kodony, [14].

Tab. 1: Odlišnosti genetického kódu a mitochondriálního genomu obratlovců

1. báze	2. báze								3. báze
	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA		UCA		UAA	Stop	UGA	Stop\Trp	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA	Ile\Met	ACA		AAA	Lys	AGA	Arg\Stop	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

V Tab. 1 je zobrazen genetický kód obratlovců černě, odlišnosti genetického kódu mitochondrie modře, start kodony mitochondrií zeleně, stop kodony mitochondrií červeně.

U kvasinek a rostlin se mitochondriální genom odlišuje více.

Většina mitochondrií má svůj původ v oocytech, z toho vyplývá, že mutace v mitochondriálním genomu (většinou jsou škodlivé) jsou přenášeny z matky na potomstvo. DNA mitochondrií podléhá mutacím častěji než DNA v jádře, možná je to zapříčiněno větším výskytem volných radikálů vzniklých při oxidativní fosforylaci, nerekombinují dostatečně rychle a poškozují mtDNA. [14]

3. Mutace

Mutace je dědičná změna v genotypu organismu. K mutacím dochází spontánně (bez vnějšího zásahu), vyskytují se s nižší frekvencí, nebo je možné je vyvolat působením mutagenů, tzv. indukované mutace. Obvykle mutace postihuje pouze malé části genetické informace buňky. Působení mutagenů na buňku případně organismus zvyšuje šanci vyvolání mutace. Mutageny mohou být biologické, chemické anebo fyzikální povahy. Změna genetické informace může mít za následek morfologické, fyziologické a biochemické změny v buňce.

Změny genetické informace způsobené mutacemi mají za následek variabilitu genomu jedinců jednoho druhu, což je základní podmínkou evoluce. Mutace tedy zapříčiňují genetickou diverzitu a různorodost genetické informace je hlavním „motorem“ evoluce a genetické diverzity.

Fyzikální mutageny působí na buňky ve formě záření. Největší podíl záření má záření ionizační a ultrafialové (UV). UV záření je pohlcováno povrchem organismu, poškozuje tedy povrchové buňky. Může být zdrojem karcinomu kůže. Záření absorbované organismem může mít původ v přirozených (radioaktivní podloží) nebo umělých zdrojích (havárie, lékařské zákroky).

Chemické mutageny jsou chemické látky poškozující genetickou informaci. Mutagenní účinky se mohou vyskytnout u metabolitu látky, která sama o sobě není genotoxická. Do chemických mutagenů se řadí např. léky (cytostatika), látky v potravě, pesticidy atd.

Zdrojem biologických mutagenů jsou především viry. Viry, konkrétně onkogenní viry dokáží poškodit buněčnou DNA a vyvolat v buňce rakovinné bujení. [9, 12]

Podle rozsahu změny je možné mutace rozčlenit na genové mutace, chromozomové mutace a genomové mutace. Modely, které jsou vypsány v následující kapitole, se zabývají genovými mutacemi, které působí změny v genotypu, ale nemění čtecí rámec.

3.1. Genová mutace

Genová mutace mění omezený úsek genetické informace, jedná se o substituci nukleotidů, inserci či delecii DNA řetězce (posunové mutace). Pokud dojde k substituci nukleotidů jedná se o změnu jedné báze, tzv. jednobodová mutace, nebo dojde k substituci většího počtu bází, vícebodová mutace. Pokud dojde k substituci nukleotidů s podobnou stavbou (změna purinu na purin nebo změna pyrimidinu na pyrimidin) jedná se o tranzici. Konkrétně se bude jednat o změny purinu na purin nebo substituci pyrimidinu za pyrimidin. Mezi puriny patří nukleotidy adenin (A) a guanin (G). Do skupiny pyrimidinů spadají cytosin (C), thymin (T) a v případě RNA uracil (U). Při transverzi se substituují nukleotidy s odlišnou stavbou. Bude tedy docházet ke změně molekuly purinu (A, G) za molekulu pyrimidinu (C, T/U) a naopak molekula pyrimidinu bude nahrazena molekulou purinu. Ve většině případů nukleotidových substitucí dochází k tranzici, transverze není tak častá. Při inserci se do sekvence vloží úsek jiné sekvence, pokud se jedná o delecii, část sekvence se oddělí. Při transverzi nebo tranzici se změní pouze kodon, kdežto inserce či delece změní i čtecí rámec. [1, 9]

3.2. Bodové mutace

Bodové mutace jsou změny, které ovlivní pouze jeden bazický pár v DNA sekvenci. Ke změnám často dochází při procesu replikace DNA nebo při vystavení se mutagenům, jako jsou například rentgenové záření či UV záření. Do bodových mutací nepatří pouze substituce, řadí se do nich i delece nebo inserce jednoho nukleotidového páru. Při translaci může také dojít k mutacím. Pokud se tak stane lze mutace rozdělit do tří skupin. První skupinou jsou synonymní mutace, do druhé skupiny patří nesynonymní mutace a do třetí skupiny řadíme nesmyslné mutace (nonsense mutations). [13]

3.3. Synonymní mutace

U synonymních mutací (označují se také jako *silent mutations – tiché mutace*) se jedná o substituci jedné báze v kodonu za jinou, ale výsledná kódovaná amino kyselina se nezmění, díky degenerovanému genetickému kódu. Když se při mutaci zamění jedna báze původního kodonu za jednu z „alternativních“ možností (viz Obr. 3), výsledná amino kyselina zůstane beze změny. To platí pouze v případě, že se změněný nukleotid v tripletu na třetí pozici. Může také dojít k změně kodonu, který kóduje odlišnou aminokyselinu. Na rozdíl od nesynonymních mutací, tato změněná aminokyselina nemění konečnou funkci výsledného polypeptidového řetězce. Při vzniku tiché mutace se změna báze neprojeví na výsledném fenotypu. [9]

3.4. Nesynonymní mutace

Nesynonymní substituce má na fenotyp mnohem větší dopad než synonymní substituce. Většinou dochází k inserci či deleci jediného nukleotidu při procesu transkripce z DNA do RNA. Prodloužením, resp. zkrácením řetězce dojde k změně zbytku čtecího rámce sekvence amino kyselin a změní se kódující kodony. Ve výsledku se změní konečný protein. Závažnost mutace záleží na místě, kde k ní došlo. Pokud k mutaci došlo blízko startu, celý kódovaný protein je změněn a tato mutace může být smrtelná.

Dále může být pozměněn pouze jeden nukleotid (bodová mutace), který ovšem ve výsledku kóduje jinou aminokyselinu. Tato změna ve většině případů nezpůsobí velké změny proteinu a není pro organismus přítěží. Pokud ovšem změna nukleotidu vede k předčasnému ukončení transkripce (kodon je zaměněn za stop kodon), protein se nevyrobí a dopad může být fatální.

Nesynonymní mutace nemusí být vždy škodlivé, některé mohou přinášet pozitivní změny. Skrze tyto změny se organismus adaptuje, a pokud se mutace vyskytne i v gametách je předána následujícím generacím. Nesynonymní substituce takto přispívají ke genetické diverzitě a evolučnímu vývoji. [13]

3.5. Nesmyslné mutace

O nesmyslné mutace se jedná v případě, pokud dojde k substituci nukleotidu a výsledkem je vznik stop kodonu na místě, kde by být neměl. Tím je zapříčiněno předčasné ukončení syntézy proteinového řetězce a pravděpodobně dojde ke ztrátě kompletně celé funkce takto postiženého proteinu.

	Bodové mutace			
	Bez mutace	synonymní	nesmyslná	nesynonymní
DNA	TTC	TTT	ATC	TCC
mRNA	AAG	AAA	UAG	AGG
protein	Lys	Lys	STOP	Arg

Lys

Thr

Arg

Obr. 3: Bodové mutace, [12]

Ukázka bodových mutací je znázorněna na Obr. 3 aminokyseliny Lysin, v prvním sloupci je AK bez mutace, ve druhém sloupci dochází k synonymní mutaci, třetí sloupec obsahuje mutaci nesmyslnou a čtvrtý a pátý sloupec obsahují mutace nesynonymní. [13]

3.6. Negativní selekce

Účelem negativní selekce je zabránit rozšíření nově vzniklých škodlivých mutací a tím zachovat kladné evoluční změny. Škodlivými mutacemi by se tyto pozitivní změny mohli ztratit. Negativní selekce je tedy velmi důležitá pro optimalizaci genetických vlastností, zaručuje že se škodlivé mutace nerozšíří v populaci. O negativní selekci se jedná v případě, že probíhající mutace v DNA nezpůsobí změnu funkčnosti vzniklého proteinového řetězce, jinak řečeno dochází k synonymní substituci. [4, 5]

3.7. Pozitivní selekce

Opakem negativní selekce je pozitivní selekce jedná se o proces, kterým do populace pronikají nové geny, zvýhodňující organismus oproti ostatním jedincům, kteří tuto genetickou změnu postrádají. Pozitivní selekce je také označována jako Darwinova selekce a je hlavním mechanismem Darwinovy evoluční teorie. Pro zjištění přítomnosti pozitivní selekce významně přispěli Yang a Neilsen. Zvýhodňující mutace v genomu se vyskytují s velmi nízkou frekvencí, oproti mutacím neutrálním anebo škodlivým mutacím, detekce prospěšných mutací je proto obtížná. Aby došlo k pozitivní selekci, musí být přítomny nesynonymní mutace ve větší míře než synonymní substituce. [5]

Pro detekci pozitivní selekce se vyvinuli dvě rozšířené metody, jak přítomnost těchto změn odhalit. První metoda je založená na analýze podstaty a četnosti odlišných alel u organismů, jedná se tedy o populační metody. Druhá metoda, analýza kodonů, porovnává hodnoty synonymních a nesynonymních změn v řetězcích kódujících aminokyseliny. [16]

3.8. Chromozomální mutace

Chromozomální mutace považujeme jako změny struktury nebo tvaru chromozomů, nikoliv však změny počtu chromozomů. Mutace na chromozomech postihují velké množství genů, a proto se mutace projevuje i na fenotypu postiženého organismu. Mutace se projevují u následující generace, která poškozené chromozomy zdědí. [13]

3.9. Genomová mutace

O genomovou mutaci se jedná v případě změny počtu jednotlivých chromozomů. Počet chromozomů se může snížit nebo zvýšit. Genomové mutace se nadále dělí na aneuploidii, polyploidii a haploidii. Aneuploidie je změna počtu chromozomů (ztráta či zmnožení). Jedná se např. o Downův syndrom, Turnerův syndrom atd. Polyploidie je znásobení celých chromozomů. Některé tkáně jsou polyploidní (některé jaterní buňky) nebo organismy (jeseter je oktaploid). Haploidní mutace vznikají snížením chromozomálních párů na polovinu. U některých živočichů je tento stav druhovým znakem (včely a trubci). U vyšších organismů je redukce latentní. [12]

Nové geny vzniklé mutacemi nemusí být nutně trvalé. Změněné nukleotidy (substitucí atd.) mohou s jistou pravděpodobností mutovat zpět na původní stav. Tato mutace se nazývá zpětná mutace. Pravděpodobnost zpětné mutace není tak vysoká jako mutace jí předcházející. [12]

4. Zarovnání sekvencí

Zarovnávání sekvencí je možné vysvětlit jako postup, při kterém porovnáváme dvě nebo více sekvencí a vyhledáváme v nich shody jednotlivých znaků nebo skupiny za sebou jdoucích znaků ve shodném pořadí. Míra podobnosti zarovnaných sekvencí reprezentuje evoluční vzdálenost mezi použitými daty.

Pokud zarovnávané stejně dlouhé sekvence, můžeme je k sobě přiřadit tak, aby začáteční znaky byly pod sebou. Poté, podle předem stanovených pravidel, určíme hodnotu zarovnání (score). Při nejjednodušším zarovnávaní počítáme pouze s hodnotami při shodě znaků (match) nebo při neshodě (mismatch).

Zarovnávání různě dlouhých sekvencí nelze provést předchozí metodou. Pro srovnávání různě dlouhých sekvencí využíváme metod lokálního nebo globálního zarovnání. Lokální zarovnání je vhodné použít, pokud jsou sekvence od sebe evolučně vzdálené. Lokální zarovnání se ukončí, když se od sebe začnou sekvence příliš odlišovat (nemusí se zarovnat celá sekvence pouze její část). Globální zarovnání se využívá u podobných sekvencí (jinak by mohlo dojít k chybám). Při použití globálního zarovnání se zarovnají celé sekvence.

Pro zvolení správné metody zarovnání lze využít bodového diagramu. Vzájemná podobnost je zobrazena tečkami na diagonále diagramu. [3, 7]

4.1. Zarovnání bez mezer

Zarovnávání sekvencí bez mezer přiřadí kratší řetězec k delšímu tak, aby se znaky sekvencí pod sebou co nejvíce shodovali. Při zarovnávaní bez mezer nelze vkládat mezery, hodnota výsledného skóre závisí pouze na volbě začátku zarovnávaní. [7]

4.2. Zarovnání s mezerami

Zarovnání s použitím mezer (gaps) vkládá mezery, které představují možné inserce nebo delece. Samozřejmě přítomnost mezer v zarovnání se musí zohlednit při výpočtu skóre. Může se také odlišit výskyt nové mezery a prodlužování již existujících mezer. [3, 7]

4.3. Vícenásobné zarovnání

Pokud zarovnávané více sekvencí najednou, jedná se o vícenásobné zarovnání. Stejně jako při zarovnávaní dvou sekvencí i při zarovnávaní více sekvencí hledáme největší míru příbuznosti. Vícenásobné zarovnání může být provedeno pomocí dynamického programování. Avšak tímto způsobem je možné zarovnat pouze malé množství sekvencí, kvůli vysoké výpočetní náročnosti. [3]

Příklad: Pokud zarovnáváme tři sekvence a každá z nich má délku tři sta znaků, je nutné provést 300^3 porovnání.

Náročnost výpočtu exponenciálně roste s počtem sekvencí.

CLUSTAL

Metoda CLUSTAL souží k vícenásobnému zarovnání sekvencí. Její výpočetní náročnost je výrazně nižší než náročnost dynamického programování. Využívá se pro zarovnávání většího počtu sekvencí. Hlavní výhodou zarovnávání pomocí CLUSTAL metody je její rychlost. Principem metody je nejprve porovnat všechny zarovnávané sekvence mezi sebou v párech a naleznout nejvíce si podobné sekvence, které následně k sobě zarovná. Výpočty provádí pomocí dynamického programování. K zarovnaným sekvencím se následně přiřazují další méně podobné sekvence nebo skupiny sekvencí až po nejodlišnější. Ve verzi CLUSTALW je možné také nastavení vah pro jednotlivé sekvence zvlášť a tím ovlivnit výsledné zarovnání. [3, 7]

Cavia	-MFMINLLLLLIIPILLAMAFLTLTER
Galeopterus	-MHTINTLSLIIPIMLAMAFLTLVER
Microtus	MVYFANILMLLVPLIAMAFLLVE-
	:. * * *: : : : :*****.*

Obr. 4: Vícenásobné zarovnání pomocí internetového nástroje ClustalO

Pomlčka (-) znázorňuje již zmiňovanou mezeru (gap). Dále pod zarovnanými sekvencemi jsou zobrazeny další znaky. Hvězdička označuje shodu aminokyselin, resp. nukleové kyseliny (pokud jsou zarovnané nukleotidové sekvence) ve všech zarovnávaných sekvencích.

5. KaKs kalkulátor

KaKs kalkulátor zaznamenává hodnoty nesynonymních (Ka) a synonymních (Ks) mutací. Jedná se o nejjednodušší počítání odlišností. Na tomto principu pracují i další složitější modely. Ovšem hodnoty Ka a Ks jsou následovně upraveny dalšími parametry (pravděpodobnost tranzice, transverze, zpětná mutace atd.). Hodnoty Ka a Ks jsou významné pro zkoumání evoluční dynamiky a pro rekonstrukční fylogenezi protein kódujících sekvencí. Do algoritmu se vkládají příbuzné organismy. Podle hodnot Ka a Ks, nebo přesněji podle jejich poměru (Ka/Ks) rozlišujeme neutrální mutaci, když se Ka rovná hodnotě Ks. Pokud je Ka menší oproti Ks jedná se o negativní selekci. Když je hodnota Ka větší, než je hodnota Ks mluvíme o pozitivní selekci. [18]

6. Evoluční modely

Modely zkoumající evoluční změny nukleotidů. Mohou porovnávat sekvence DNA shodné délky a následně určovat jejich míru shodnosti. Porovnáváme sekvence, u kterých předpokládáme, že se jedna sekvence vyvinula z druhé porovnávané, nebo při porovnávání sekvencí vyvinutých ze společného předka. Jako výsledek získáme počet všech substitucí nukleotidů, nebo takzvanou evoluční vzdálenost mezi porovnávanými sekvencemi. Ze vzdálenosti se dále určí čas, kdy se z jedné sekvence vyvinula nová, nebo před jakou dobou se oddělili od společného předka. [6, 17]

6.1. Základní evoluční modely

Evoluční modely je možné rozdělit na dvě skupiny. Do první skupiny by se řadily modely využívající k výpočtům aproximační metody. Do druhé skupiny by spadaly modely počítající pomocí metody maximum likelihood.

Aproximační modely provádí výpočty v následujících krocích. Za prvé naleznou místa výskytu synonymních a nesynonymních mutací, následně spočítají počty synonymních a nesynonymních mutací. Na konec provedou korekci pro vícenásobnou substituci.

Oproti aproximačním metodám, metody maximum likelihood spojují evoluční metody a pravděpodobnostní metody a díky tomu jsou schopny vypočítat všechny kroky aproximačních modelů v jediném kroku. [6]

6.2. Proporcionální vzdálenost sekvencí

Proporcionální vzdálenost sekvencí je jeden z nejjednodušších způsobů, kterými lze vyjádřit evoluční vzdálenost dvou sekvencí. Stačí naléznout všechny pozice, na kterých jsou sekvence odlišné a výsledek vydělit délkou sekvence. Získaná hodnota se nazývá p-distance. Při analýze výpočet nehledí, zdali došlo k mutacím, které nejsou zaznamenatelné v konečném produktu, jako např.: zpětné mutace ($A \rightarrow G$; $G \rightarrow A$), nebo více mutací na stejné pozici ($A \rightarrow T$; $T \rightarrow G$). [7]

6.3. Neutrální model

Neutrální model byl představen panem Kimurou v šedesátých letech. V neutrálním modelu je naprostá většina mutací rozdělena do dvou skupin. První skupina mutací vstupující do modelu jsou neutrální mutace, které se staly součástí genomu působením genetického driftu. Většina změn nukleotidů mezi dvěma organismy jsou změny tohoto typu. Do druhé skupiny patří škodlivé mutace, které se objevují náhodně a jsou eliminovány přirozeným vývojem. Ačkoliv se jedná o negativní změny v genomu, tyto rozdíly přispívají k diverzitě populace. [16]

6.4. Jukes-Cantor model

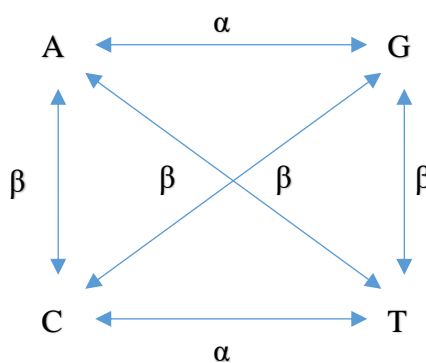
Jukes-Cantor model (JC69), vytvořený pány Thomasem Jukesem a Charlesem Cantorem v roce 1969, je jeden z prvních substitučních modelů používaných k odhadnutí evoluční vzdálenosti. Vychází z předpokladu, že všechny substituce jsou zcela nezávislé, vyskytují se náhodně se stejnou pravděpodobností, k mutaci může dojít na jakékoliv pozici v sekvenci a nedochází k inzercím ani delecím. Pro výpočet využívá jediné proměnné. Model je popsán následující rovnicí [6, 11]:

$$D = -\frac{3}{4} \ln\left(1 - \frac{4}{3}\mu\right) \quad (6.1)$$

Kde D je vypočítaná evoluční vzdálenost mezi sekvencemi a μ je počet neshodujících se nukleotidů mezi zarovnanými sekvencemi. [3, 6, 17]

6.5. Kimura model

Kimurův model (K80) byl vytvořen panem Kimurou v roce 1980 na základě rozdílu četnosti mezi tranzicemi a transverzí v mitochondriálním genomu zveřejněného ve stejném roce. Nepředpokládá, že všechny nukleotidy mezi sebou mutují se stejnou pravděpodobností a také nepoužívá jednu proměnnou pro výpočet, na rozdíl od modelu JC69. Stejně jako předchozí model předpokládá stejnou frekvenci výskytu všech nukleotidů (25%). Nukleotidovou substituci můžeme rozdělit na dvě podskupiny: tranzici a transverzi. Proto se tomuto modelu také říká: Kimura two-parametr model (K2P). [6, 17]



Obr. 5: Kimurův dvouparametrový model

Na obrázku Obr. 5. je zobrazena odlišnost mezi tranzicemi (α) a transverzí (β). Je patrné, že každý nukleotid může podlehnout jedenkrát tranzici s pravděpodobností α anebo jsou dvě možnosti transverze s četností β . [6, 17]

Model je popsán rovnicí:

$$K = -\frac{1}{2} \ln((1 - 2\alpha - \beta) \cdot \sqrt{1 - 2\beta}) \quad (6.2)$$

Kde α je pravděpodobnost tranzice a β je pravděpodobnost transverze.

Tab. 2: Míra substitucí nukleotidů, K80

Kimura – tabulka četnosti tranzice a transverze				
	A	T	C	G
A	-	β	β	α
T	β	-	α	β
C	β	α	-	β
G	α	β	β	-

6.6. Felsenstein model

Felsensteinův model (F81), vyvinutý v roce 1981 Josephem Felsensteinem, pracuje na základě modelu Jukes-Cantor. Odklání se od něj při počítání četnosti substitucí. Tento model nepočítá se stejnými frekvencemi substitucí nukleotidů, jako JC69, protože některé nukleotidy se v sekvencích vyskytují ve větší míře než jiné, a proto u těchto nukleotidů dochází k častějším substitucím a naopak. Míry substitucí jsou znázorněny v Tab. 3.

Tab. 3: Míry substitucí nukleotidů, F81

Felsenstein – tabulka pravděpodobnosti tranzice a transverze				
	A	T	C	G
A	-	π_T	π_C	π_G
T	π_A	-	π_C	π_G
C	π_A	π_T	-	π_G
G	π_A	π_T	π_C	-

Evoluční vzdálenost (d) modelu F81 lze vypočítat následovně:

$$d = -(1 - (\pi A^2 + \pi C^2 + \pi G^2 + \pi T^2)) * \log(1 - pDistanc / (1 - \pi A^2 + \pi C^2 + \pi G^2 + \pi T^2)) \quad (6.3)$$

Kde π_A , π_G , π_C a π_T jsou míry četností příslušných nukleotidů v sekvenci a pDist je hodnota genetické vzdálenosti. [6]

6.7. Maximum likelihood model

Metodu maximum likelihood (ML) k analýze molekulových sekvencí začal využívat Jerzy Neyman v roce 1971. Uvedení do praxe umožnili výkonnější počítače, protože metoda ML je při aplikaci např. biologických dat poměrně výpočetně náročná. Z toho vyplývá, že není možné analyzovat velké množství sekvencí (v porovnání s jednoduššími modely) v krátkém čase. Aproximací výsledků je možné urychlit výpočty.

Hlavním kladem metody ML je aplikace velmi dobře propracovaných evolučních modelů na analyzovaná data čímž získáme přesnější a věrohodnější výsledky. [3, 6]

7. Řešení problematiky v programu R

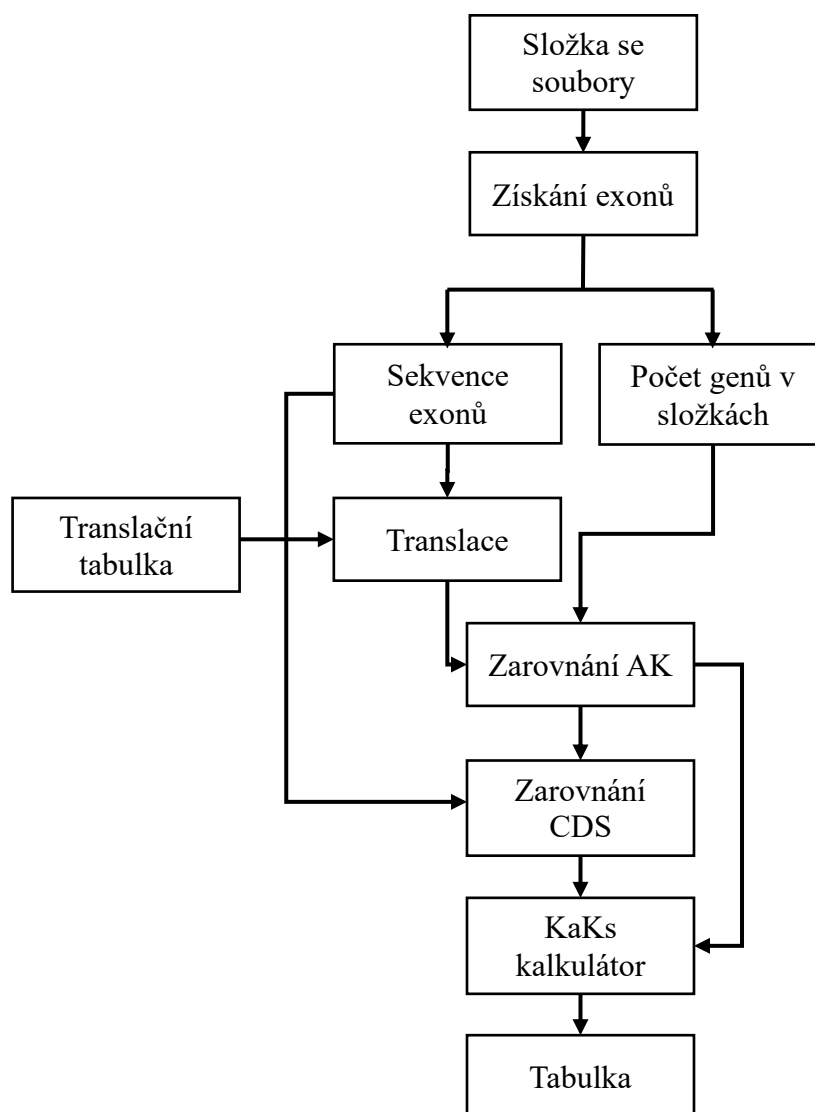
Pro získání sekvencí a následné jejich zpracování je zapotřebí mít v programovém prostředí nainstalovány následující knihovny pro zajištění funkčnosti. Jedná se o knihovny *genbankr*, *Biostrings*, *msa* a *xlsx*. Knihovna *genbankr* získává ze souborů formátu *genbank* nukleotidové sekvence a další potřebné k získání exonů. Knihovna *Biostrings* obsahuje funkce ke zpracování sekvencí (translace atd.). Vícenásobné zarovnání sekvencí umožňuje knihovna *msa*. Přepsání výsledků do tabulky formátu *xls* umožňuje knihovna *xlsx*.

Klíčovým problémem bylo zpracování mitochondriálních DNA obratlovců v prostředí programu R Studio a následovně porovnat výsledky. Nejprve bylo nutné získat nukleotidový řetězec ze souboru ve formátu *genbank* od každého organismu a poté z něj vyjmout kódující úseky (exony) jednotlivých genů. Po získání exonů bylo nutné provést jejich translaci podle translační tabulky pro mitochondrie obratlovců na aminokyselinové (AK) sekvence. Převedené AK sekvence je nutné zarovnat pomocí *multi sequence alignment (msa)*, knihovna v R Studiu pracuje na stejném principu jako webový nástroj ClustalW. Podle jednotlivých zarovnaných AK sekvencí se zarovnají i nukleotidové sekvence. Takto zarovnané sekvence se porovnávají a odlišné páry nukleotidů, resp. AK se zaznamenávají do hodnot synonymních a nesynonymních mutací. Přesný postup bude rozebrán v následujících kapitolách. Výsledkem této práce je tabulka obsahující názvy dvojic organismů a genu, počet nesynonymních a synonymních mutací a jejich poměr, který vypovídá o evoluční dynamice porovnávaných organismů.

Schéma na Obr. 6 ilustruje tok dat mezi vytvořenými funkcemi.

Funkce *getCDS*, znázorněna v Obr. 6, slouží k získání úseků mtDNA, které kódují jednotlivé geny. Vstupní proměnnou pro tuto funkci je umístění souborů s příponou *.gbk*, které se mají zpracovat. Je nutné mít v R Studiu nainstalovány knihovny *genbankr* a *Biostrings*. Je nutné vkládat sekvence se stejným počtem genů (v tomto případě měli všechny sekvence 13 genů). Funkce nalezne všechny soubory s příponou *gbk* ve složce určené vloženou cestou. K získání názvů všech souborů je použit příkaz *list.files(pattern = '*.gbk')*. Po nahrání souboru se do proměnných uloží potřebná data např.: celá nukleotidová sekvence, startovní nukleotidy jednotlivých genů, délku sekvence pro každý gen atd. Výstupní proměnnou jsou tedy jednotlivé sekvence genů všech vstupujících organismů s jejich názvy a počet genů v sekvencích.

Pomocí funkce *translation* převedeme získané nukleotidové sekvence na AK sekvence. Funkce je ve schématu Obr. 6. Vstupy pro tuto funkci jsou translační tabulka (1 = DNA obratlovců, 2 = mtDNA obratlovců atd.) a nukleotidový řetězec. Pomocí funkce *getGeneticCode* získáme příslušnou translační tabulku. Ke každé sekvenci se přiřadí odpovídající název. Výstupem funkce *translace* je AK sekvence.



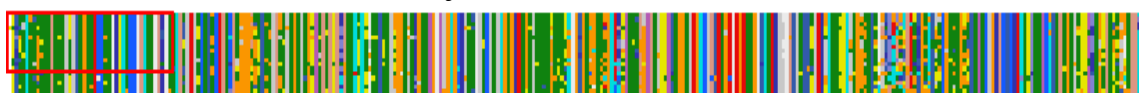
Obr. 6: Blokové schéma vytvořených funkcí

Zarovnání AK sekvencí (Obr. 7) obstarává funkce *zarovnani_AA_sek*, která využívá knihovny *msa*. Je tedy nutné si tuto knihovnu do programu R Studio nainstalovat stejně jako předešlé knihovny *genbankr* a *Biostrings*. Zarovnáno je devatenáct sekvencí genu ND1. Vstupními hodnotami jsou sekvence aminokyselin získané předešlou funkcí *translace* a počet genů obsažen v porovnávaných souborech. Po načtení knihovny *msa*, se oddělí názvy organismů od názvu genů. Tento krok je potřeba pro identifikaci stejných

genů odlišných organismů. Původní záměr pro vyhledání indexů genů byl využit funkce *grep*, avšak při vyhledávání umístění genu ND4, funkce vzhledala i pozice genu ND4L. Sekvence se zarovnají pomocí funkce *msa* mnohonásobným zarovnáním, podobně jako při využití internetového nástroje ClustalO. Zarovnávání je nastavenou na defaultní hodnoty. Výstupem funkce *zarovnani_AA_sek* jsou vícenásobně zarovnané sekvence aminokyselin.

Obr. 7: Vícenásobné zarovnání aminokyselin

Po zarovnání AK sekvencí je nutné zarovnat nukleotidové řetězce, k tomuto účelu



souží funkce *zarovnani_CDS*. Do funkce vstupují zarovnané AK sekvence z funkce *zarovnani_AA_sek* a nezarovnané nukleotidové sekvence získané funkcí *getCDS*. Znovu bylo záměrem využít funkce *grep*, ale docházelo ke stejnému problému, a proto od něj bylo upuštěno. Zarovnávají se pouze sekvence CDS, které mají totožným název jako sekvence AK (název organismu i genu jsou shodné), který porovnáváme podmínkou *if*. Zarovnání probíhá následovně: AK sekvence je kontrolována prvek po prvku. Pokud je přítomna aminokyselina, do pomocné proměnné se uloží triplet ze sekvence CDS se stejným názvem. Pozice tripletu odpovídá pozici aminokyseliny. Pokud je ovšem namísto aminokyseliny nalezena mezera (-), do zarovnané nukleotidové sekvence není uložen odpovídající triplet, ale jsou vloženy tři mezery (---). Po uložení tripletu nebo mezer, je kontrolována následující aminokyselina až do konce řetězce, poté se přechází na novou sekvenci. Výstupem z funkce *zarovnani_CDS* jsou zarovnané nukleotidové sekvence podle odpovídajících AK sekvencí.

Funkce *KaKs_kalkulacka* porovnává zarovnané sekvence nukleotidů a aminokyselin. Výstupem jsou hodnoty synonymních mutací (Ks), nesynonymních mutací (Ka), jejich poměr Ka/Ks, Jukes-Cantor model a Kimmura model. Vstupem funkce jsou zarovnané NC a AC sekvence.

Funkce *JC* a Kimura-2P.R vstupují do fce *KaKs_kalkulacka*. Vstupní hodnoty jsou dosazeny do rovnic

$$(D = -\frac{3}{4} \ln(1 - 4/3\mu)) \quad (6.1)$$

Případně

$$K = -\frac{1}{2} \ln((1 - 2\alpha - \beta) \cdot \sqrt{1 - 2\beta}) \quad (6.2).$$

Výsledky rovnic jsou výstupními hodnotami těchto funkcí.

Pro výpočet poměru synonymních a nesynonymních mutací vstupují hodnoty, konkrétně páry kodonů dvou porovnávaných sekvencí, do pomocných funkcí. Jedná se o funkce *NSites*, *KaKsRatio* a *dNdSRatio*.

První pomocnou funkcí vstupující do *KaKs kalkulátoru* je funkce *NSites*, vstupem je kodon první z dvojce porovnávaných řetězců a příslušná translační tabulka ve funkci *translace*. Vývojový diagram funkce je znázorněn na *Obr. 8*. Kodon se stává z třech nukleotidů (A, G, C, T), může tedy dojít k devíti různým mutacím v celém tripletu. Funkce počítá očekávaný počet nesynonymních pozic pro každý kodon. Ukázka názorného výpočtu pro AK methionin (Met, ATG) je znázorněna v .

Tab. 4: Příklad výpočtu funkce *NSites* pro AK Methionin (Met, ATG)

Pozice v kodonu	substituce	Výsledný triplet	Výsledná aminokyselina	Typ mutace	Počet nesynonymních pozic (n)
1.	A→G	GTG	Val	nesynonymní	1/3
	A→T	TTG	Leu	nesynonymní	1/3
	A→C	CTG	Leu	nesynonymní	1/3
2.	T→A	AAG	Lys	nesynonymní	1/3
	T→G	AGG	stop	nesmyslná	1/3
	T→C	ACG	Thr	nesynonymní	1/3
3.	G→A	ATA	Met	synonymní	0
	G→	ATT	Ile	nesynonymní	1/3
	G→	ATC	Ile	nesynonymní	1/3

Výsledná hodnota nesynonymních pozic v aminokyselině se získá následujícím vzorcem. Že se jedná o nesynonymní mutaci zjistíme translací nově získaného triplet (k tomu je třeba vstupní proměnná: translační tabulka).

$$n = \sum_{i=1}^3 f_i \quad (7.1)$$

Kde n je celkový počet nesynonymních pozic a f_i je hodnota nesynonymního místa pro i -tou pozici. A hodnotu synonymních pozic (s) udává vzorec:

$$s = (3 - n) \quad (7.2)$$

Výsledek výše uvedeného příkladu by vypadal takto:

- $n = (1/3 + 1/3 + 1/3) + (1/3 + 1/3 + 1/3) + (0 + 1/3 + 1/3) = 2,667$
- $s = 3 - 2,667 = 0,333$

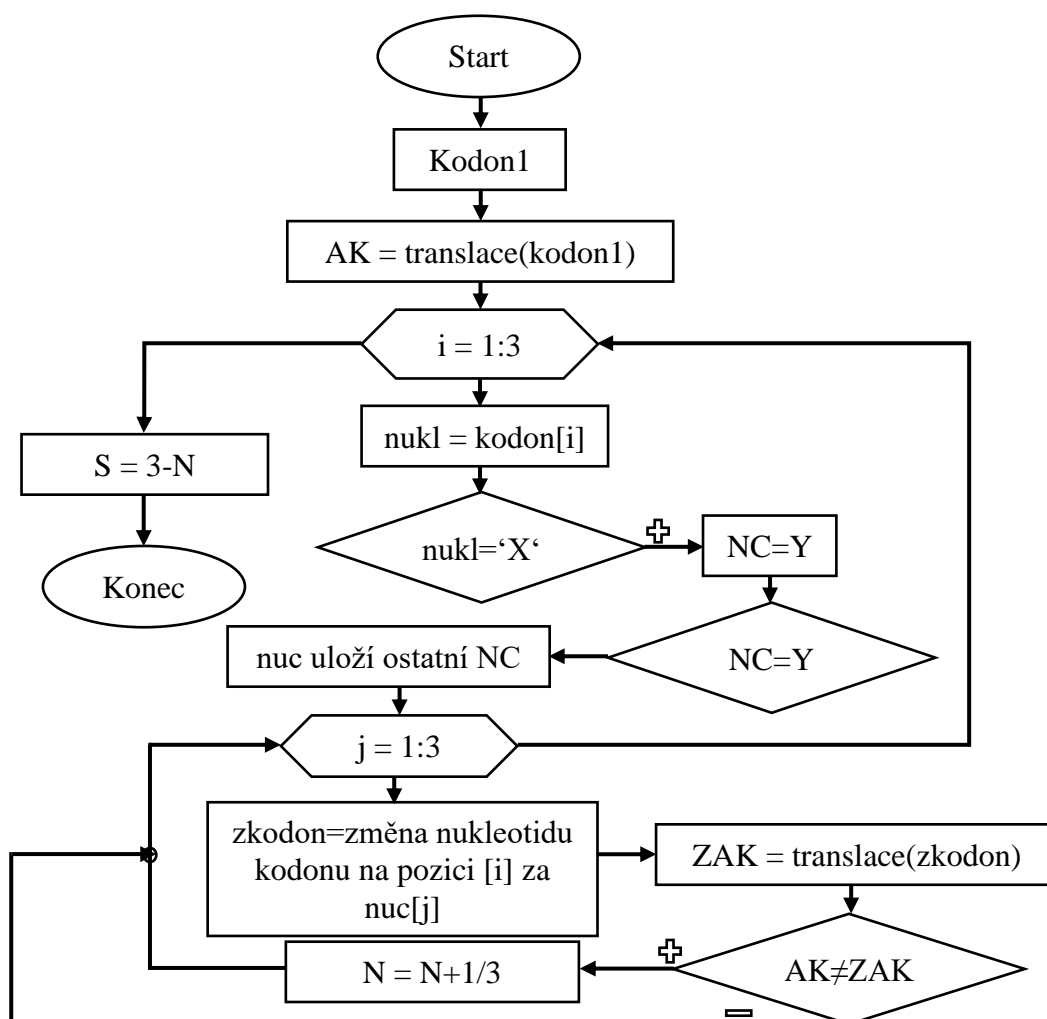
Počet nesynonymních pozic pro methionin ATG je 2,667 zatímco počet synonymních pozic je 0,333. Hodnota pro místa nesynonymních a synonymních mutací pro celou sekvenci se vypočítá podle následujících rovnic

$$N = \sum_{i=1}^r n_i \quad (7.3)$$

$$S = (3 \cdot r - N) \quad (7.4)$$

Kde N je celkový počet nesynonymních pozic pro celou sekvenci, r počet kodonů v sekvenci, n_i je počet nesynonymních pozic i -tého kodonu a S je počet synonymních pozic pro celou sekvenci.

Původně se ve funkci neměla vyskytovat první podmínka, ale z neznámého důvodu program R hlásil chybu, kterou se nepodařilo vyřešit, a proto bylo zavedeno náhradní řešení s prvními podmínkami.



Obr. 8: Vývojový diagram funkce *NSites*

Na Obr. 8 zobrazena je část pouze pro znázornění postupu průběhu výpočtu, hodnota X zde zastupuje nukleotid (A, G, C, T), který vstupuje do podmínky, pokud je splněna do hodnoty NC se uloží Y ($A=1, G=2, \dots$). Pokud by do první podmínky vstoupil adenin (A), NC by se rovnalo 1 a do nuc by se uložili G, C, T, zbytek probíhá identicky u všech možností.

Další pomocná funkce vstupující do KaKs kalkulátoru je *KaKsRatio*, která porovnává vstupující kodony srovnávaných řetězců a získává počet nesynonymních a

nesynonymních substitucí. Vstupem jsou tedy již zmíněné kodony obou porovnávaných sekvencí a také translační tabulka stejně jako u předešlé funkce. Pokud se v tripletech vyskytuje pouze jedna substituce výpočet je jednoduchý (mutace může být pouze nesynonymní anebo synonymní). Pokud došlo ke dvěma nebo třem substitucím a výsledná mutace je nesynonymní, výpočet je složitější. Záleží na pořadí, v jakém mutace nastaly, protože ne všechny přítomné substituce musejí být nutně nesynonymní. Musejí se tedy zvážit všechny možnosti jaké mohli nastat. V případě dvou mutací jsou možnosti pouze dvě, v případě třech substitucí je cest šest. Zde je názorný příklad:

Kodon 1: Arginin, Arg, CGC; kodon 2: Tryptofan, Trp, TGG

1. cesta: CGC (Arg) → CGG (Arg) → TGG (Trp)
1 nesynonymní mutace, 1 synonymní mutace
2. cesta: CGC (Arg) → TGC (Cys) → TGG (Trp)
2 nesynonymní mutace

Po analýze obou možností je počet zpozorovaných nesynonymních mutací $n_d = 1,5$ a počet synonymních mutací $s_d = 0,5$ (součet hodnot bude roven počtu možností).

Pokud porovnáваме celé sekvence, celkový součet n_d a s_d se spočítá pomocí rovnic:

$$N_d = \sum_{i=1}^r n_{di} \quad (7.5)$$

$$S_d = \sum_{i=1}^r s_{di} \quad (7.6)$$

Kde N_d je celkový součet nesynonymních mutací v celé sekvenci, r je počet kodonů v celé sekvenci, n_{di} je počet nesynonymních mutací v i -tém kodonu, S_d je celkový počet synonymních mutací v celé sekvenci a s_{di} je počet synonymních mutací v i -tém kodonu. Výstupem funkce *KaKsRatio* je celkový počet nesynonymních a synonymních substitucí.

Poslední pomocnou funkcí vstupující do *KaKs kalkulátoru* je *dNdSRatio*, která vypočítává poměr synonymních a nesynonymních mutací. Vstupními proměnnými jsou celkový počet nesynonymních mutací (N_d), celkový počet synonymních mutací (S_d), počet nesynonymních pozic (N) a počet synonymních pozic (S). Výpočet je realizován následujícími rovnicemi:

$$p_N = \frac{N_d}{N} \quad (7.7)$$

$$p_S = \frac{S_d}{S} \quad (7.8)$$

Kde p_N je poměr nesynonymních rozdílů a p_S je poměr synonymních rozdílů.

Následuje odhadnutí počtu nesynonymních substitucí (d_n) a synonymních substitucí (d_s) na pozici podle následujících rovnic:

$$d_N = -\frac{3}{4} \cdot \ln\left(1 - \frac{4 \cdot p_N}{3}\right) \quad (7.9)$$

$$d_S = -\frac{3}{4} \cdot \ln\left(1 - \frac{4 \cdot p_S}{3}\right) \quad (7.10)$$

Hodnoty d_N a d_S dále vstupují do rovnice pro výpočet poměru nesynonymních a synonymních mutací. Rovnice je následující:

$$dN/dS = \frac{d_N}{d_S} \quad (7.11)$$

Výstupem funkce *dNdSRatio* je poměr nesynonymních a synonymních substitucí v porovnávaných sekvencích.

Výsledky, získané funkcí *KaKs_kalkulacka*, zaznamenaná do excelové tabulky funkce *tabulka* (Obr. 6). Vstupem je list hodnot obsahující názvy porovnávaných organismů, hodnoty K_a , K_s jejich poměr, hodnoty modelů JC69 a K80.

Pomocí funkce *Posuvne_okno* je sekvence analyzována v posuvném okně. Parametry okna se zadávají, jako vstupující proměnné. Cílem posuvného okna je odhalit synonymní substituci, vyskytující s větší koncentrací na kratších úsecích.

Schéma vytvořených funkcí je zobrazeno na Obr. 6.

8. Výsledky

Výstupem této práce je tabulka s porovnávanými geny všech vložených organismů, společně s hodnotami poměru K_aK_s , a hodnot získaných modelem Jukes-Cantor a také soubor skriptů, kterými byly výsledky získány.

V tabulce je zobrazen část výsledků získaných funkcí *KaKs_kalkulátor*. Do funkce bylo vkládáno dvacet organismů. Organismy vložené do funkce jsou drobní savci např. *Cavia porcellus* (morče domácí), *Oryctolagus cuniculus* (králík divoký), *Glis glis* (plh velký) a další. Hodnoty poměru K_aK_s ku délce sekvence jsou nižší oproti hodnotám získaným modelem JC69 z důvodu, že model Jukes-Cantor počítá s jistou pravděpodobností, že u některých změn došlo k reverzní mutaci. To má za následek, že poměr K_aK_s tyto mutace není schopen odhalit, ale model JC69 je do svého výsledku započítává. Z hodnot K_aK_s je zřetelné, že u převážné většiny zvolených organismů dochází k negativní selekci. Pouze u dvojice *Sciurus vulgaris* a *Oryctolagus cuniculus* je hodnota modelu JC69 větší než 1 a to značí přítomnost pozitivní selekce. Při srovnávání sekvencí náležících k *Oryctolagus cuniculus* s řetězcí ostatních organismů jsou hodnoty získané modelem Jukes-Cantor v naprosté většině mírně vyšší než jedna. Hodnoty pozitivní selekce výrazně nepřevyšují hodnotu jedna, z čeho usuzují, že se nejedná o významný výsledek. Jak již bylo řečeno, pozitivní selekce má velmi nízkou frekvenci výskytu. Tuto skutečnost potvrzují i získané výsledky.

Tab. 5: Výsledné hodnoty funkce *KaKs kalkulátor* pro gen ND1

Referenční organismus	Porovnávaný organismus	Poměr KaKs	Míra substituce JC69
<i>Oryctolagus cuniculus</i>	<i>Lepus europaeus</i>	0,029549	0,193173
	<i>Ochotona collaris</i>	0,063418	0,257669
	<i>Microtus kikuchii</i>	0,108927	0,331461
	<i>Tupaia belangeri</i>	0,089691	0,267988
	<i>Galeopterus variegatus</i>	0,081912	0,317061
	<i>Pteropus scapulatus</i>	0,092726	0,310748
	<i>Echinops telfairi</i>	0,066547	0,351102
	<i>Macroscelides proboscideus</i>	0,087349	0,362802
	<i>Cavia porcellus</i>	0,080097	0,347793
	<i>Thryonomys swinderianus</i>	0,102898	0,376400
	<i>Glis glis</i>	0,068318	0,364488
	<i>Sciurus vulgaris</i>	0,056107	0,323427

Pokud se poměr synonymních a nesynonymních mutací ku délce řetězce počítá pro celou délku zarovnané sekvence, je možné že celkový získaný výsledek označuje selekci porovnávaných živočichů za negativní. Tento výsledek však nemusí být relevantní. V sekvenci se mohou na krátkém úseku vyskytovat nesynonymní mutace představující pozitivní selekci. Hodnota shluku mutací se v celkové délce výrazně delší sekvence ztratí. Pro odhalení významných hodnot ukrývajících se na malých úsecích dlouhého řetězce je možné využít porovnávání sekvencí pomocí takzvaného plovoucího okna. Hodnoty získané oběma metodami nezískali jedinou hodnotu větší než jedna, z toho vyplývá že ve všech analyzovaných kombinacích dochází k purifikační, negativní, selekci. Sekvence se podrobí analýze pomocí plovoucího okna, aby se zjistili, zda v sekvencích nejsou skryté pozitivní selekce.

Plovoucí okno aplikované na zarovnané sekvence mělo parametry nastaveny na následující hodnoty: šířka okna byla nastavena na šířku 99 a po proběhnutí výpočtů se okno posunulo o devět prvků. Získané hodnoty byly následovně vyneseny do grafů. Příklad je znázorněn na Obr. 9. Hodnoty v grafu na ose x znázorňují délku porovnávaných sekvencí, tedy počet nukleotidů, resp. aminokyselin. Na ose y je uveden poměr substitucí vůči délce sekvence. Pokud hodnoty na ose y překračují hodnotu jedna, je pravděpodobné, že na tomto místě (hodnotě na ose x) došlo k pozitivní selekci. Když je hodnota nižší než jedna, jedná se o negativní selekci.

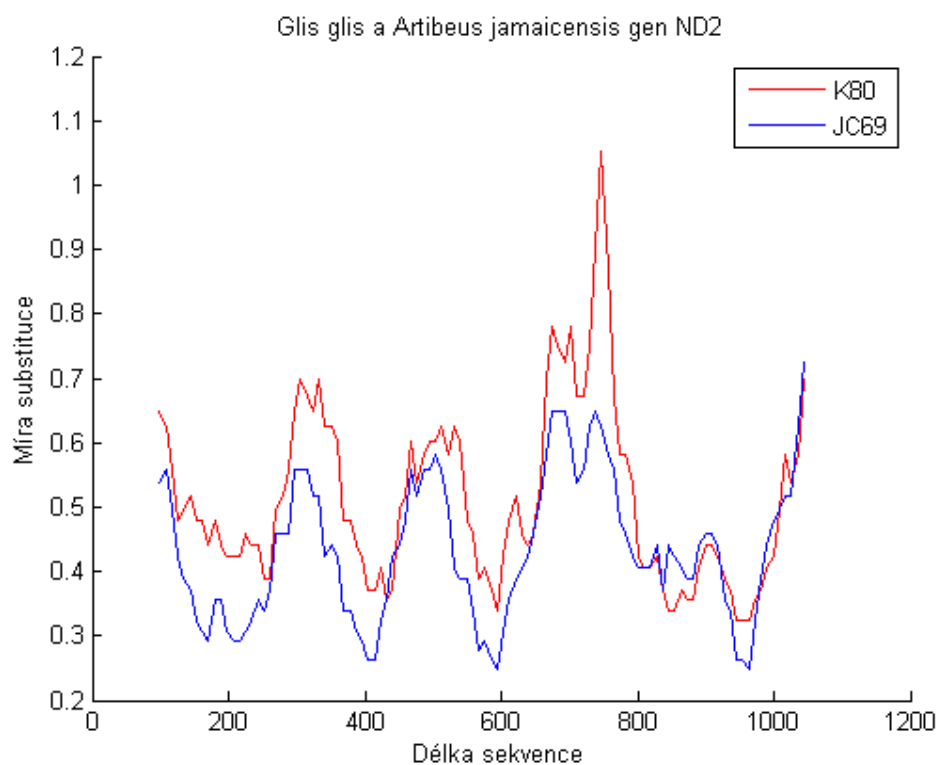
Některé výstupní hodnoty nemají číselnou hodnotu a jsou značeny NaN (not a number). Dochází tomu pouze v případech výsledků poměru KaKs. Těchto výsledků je

možné dosáhnou vícero možnostmi. Jednou možností je, že jsou porovnávané sekvence až příliš evolučně vzdálené. Další možnou příčinou je příliš krátká délka sekvencí, nebo přítomnost inserce či delecce (na změnu čtecího rámce je metoda velmi citlivá). Důsledkem je poměr synonymních substitucí v sekvencích větší než $\frac{3}{4}$, což má za následek vstoupení záporné hodnoty do logaritmu a výstup NaN hodnoty. V je uvedeno pár příkladů.

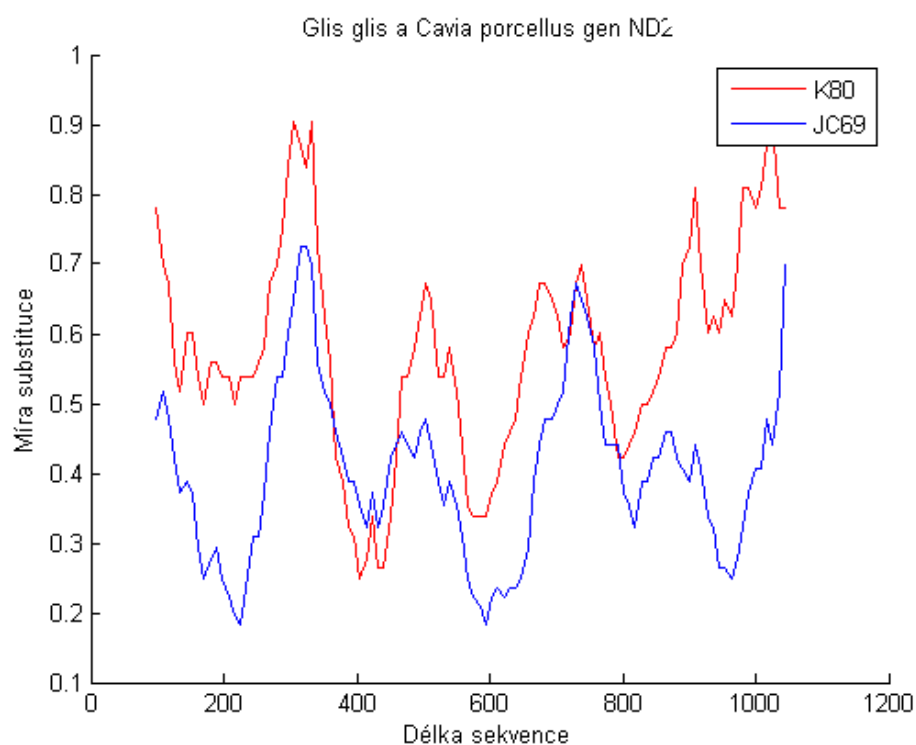
Tab. 6: Hodnoty NaN

Gen	Referenční organismus	Porovnávaný organismum	Poměr KaKs	JC69
ND3	Ochotona collaris	Sciurus vulgaris	NaN	0,376444
	Ochotona collaris	Echinops telfairi	NaN	0,446030
ND4L	Galeopterus variegatus	Oryctolagus cuniculus	NaN	0,405330
	Galeopterus variegatus	Lepus europaeus	NaN	0,405330
	Thryonomys swinderianus	Tupaia belangeri	NaN	0,517339
	Thryonomys swinderianus	Galeopterus variegatus	NaN	0,478115

Hodnoty NaN byli získány pouze při počítání poměru nesynonymních a synonymních substitucí, protože do JC69 vstupují hodnoty počtu substitucí celkem. Celkový počet nečíselných hodnot pro analyzované sekvence je třicet pět, z toho jedenáct přísluší genům ND4L. Je tedy možné, že u genu ND4L může častěji docházet například k delecím, resp. insercím, nebo se gen u různých organismů značně odlišuje.



Obr. 9: Míra substituce nukleotidů pro gen ND2 organismů *Glis glis* a *Artibeus jamaicensis*



Obr. 10: Míra substituce nukleotidů pro gen ND2 organismů *Glis glis* a *Cavia porcellus*

Při hledání substitucí mezi organismy *Glis glis* a *Artibeus jamaicensis* (porovnávaný gen ND2) výsledná míra substituce bez použití plovoucího okna je rovna 0,51 pro JC69 a 0,53 pro model K80. Pokud je použito plovoucí okno () míra substituce se pohybuje okolo délky 700 nad hranicí pozitivní selekce. Podle výsledků je tedy pravděpodobné očekávat přítomnost pozitivní selekce u AK na odpovídající pozici a jejím okolí.

Při porovnávání sekvencí organismů *Glis glis* a *Cavia porcellus* zobrazené na Obr. 10 (gen ND2) vyšla hodnota míry substituce nukleotidů získaná modelem JC69 0,44 a hodnota získaná modelem K80 0,45. Uvedené hodnoty byly získány při porovnávání celých sekvencí. Při použití plovoucího okna se hodnoty až k 0,9, ale hodnotu míry substituce větší než 1 nedosahují, z toho můžeme usoudit, že není přítomna pozitivní selekce.

Pro porovnání výsledků je využit program PAML. Vstupem pro PAML jsou zarovnané sekvence a strom evolučních vzdáleností, který je možné získat například v programovém prostředí MATLAB příkazem *getnewickstr*. Výsledkem je textový dokument obsahující výsledky analýzy. Ze všech hodnot hledáme Bayes Empirical Bayes (BEB), kde jsou znázorněny hodnoty s vyššími výsledky poměru KaKs. Vložené sekvence PAML vyhodnotil následovně:

Tab. 7: Hodnoty analyzovaných sekvencí v programu PAML

Organismus/gen	Pozice v sekvenci	Hodnota pozitivní selekce	Hodnota po zprůměrování
<i>Oryctolagus cuniculus</i> /ND1	168 T	0.504	1.544 +- 1.329
	251 T	0.516	1.480 +- 1.135
<i>Glis Glis</i> /ND2	6 A	0.588	1.162 +- 0.419
	7 T	0.719	1.284 +- 0.353
	10 F	0.537	1.085 +- 0.451
	15 I	0.690	1.349 +- 0.000
	87 F	0.655	1.372 +- 0.000
	90 T	0.513	1.108 +- 0.409
	93 F	0.597	1.172 +- 0.413
	103 T	0.580	1.192 +- 0.330
	151 P	0.587	1.165 +- 0.415
	156 T	1.136**	2.508 +- 0.000
	221 L	0.550	1.142 +- 0.409
	223 N	0.608	1.186 +- 0.395
	233 L	0.576	1.174 +- 0.374
	281 S	0.713	1.279 +- 0.357
	324 I	0.669	1.241 +- 0.379
	336 L	1.014**	2.382 +- 0.000

Hodnoty označené * jsou z 95 % pozitivní selekcí a hodnoty označené ** jsou z 99 % pozitivní selekcí. V je zobrazena část výsledků získaných programem PAML. Naprostá většina výsledků nepodléhá pozitivní selekci, ale v tomto příkladu jsou dvě hodnoty, které mají významnou hodnotu a s pravděpodobností 99 % u nich došlo k pozitivní selekci. Konkrétně se jedná o aminokyselinu Thr na pozici 156, hodnota pozitivní selekce je 1,136 a aminokyselina Leu na pozici 336 s hodnotou pozitivní selekce 1,014. Hodnoty jsou získané pro organismus *Gils Gils* (plch velký).

9. Závěr

Cílem bakalářské práce bylo analyzovat geny mitochondriální DNA a naleznout v nich přítomnost pozitivní či negativní selekce. Bakalářská práce je rozdělena na dvě části. Teoretická část se věnuje základům genetiky mitochondrií, mutacím a evolučním modelům. Druhá část se zabývá samotným procesem tvorby funkcí pro výpočet potřebných veličin. Výsledky jsou porovnávány s hodnotami získanými programem PAML.

Pro zpracování dat byly vytvořeny pomocné funkce v programovém prostředí R. V prvním kroku je zapotřebí získat úseky mtDNA funkcí *getCDS*, získané řetězce se převedou na sekvence AK pomocí translace. Před samotným porovnáváním sekvencí je nutné, aby byly mezi sebou zarovnané. K tomuto účely slouží funkce *zarovnání AA sek* pro zarovnání aminokyselinových řetězců a pro zarovnání řetězců nukleotidů je funkce *zarovnání CDS*. Výsledkem je výstup funkce *KaKs kalkulátor*, který obsahuje hodnoty poměru pozitivní a negativní selekce, a dále výsledky modelů JC69 a K80.

Celkem bylo analyzováno devatenáct organismů, konkrétně se jedná o drobné savce, převážně hlodavce. Geny ND2 a ATP8 dosahují nejvyšších hodnot (cca od 0,45 až 0,8) míry substituce. Z toho vyplývá, že v těchto genech dochází k nesynonymním mutacím častěji než v ostatních genech. Avšak žádný získaný výsledek nepřesahuje hodnotu jedna, to znamená že všechny porovnávané sekvence podléhají negativní selekci. Jedny z nejvyšších hodnot dosahuje živočich *Glis glis* při porovnávání s ostatními živočichy. Hodnoty jsou zaznamenány v Tab. 8 a Tab. 9. Získané výsledky potvrzují obecně známá fakta, že se pozitivní selekce vyskytuje s velmi malou frekvencí.

Program PAML slouží k analýze zarovnaných sekvencí NK, je citlivější než vytvořené funkce a pomocí komplexnějších výpočtů získává pravdivější výsledky. Po vložení dat odhalil pozitivní selekci v organismu *Glis Glis* (Plch Velký), kterou vytvořené funkce nezaznamenali. Konkrétně v genech ND2 a ATP8. Pozice pozitivní selekce je zobrazena v

Příloha 3

Tab. 10. Z toho vyplývá, že do genomu tohoto živočicha pronikly zvýhodňující geny. Hodnoty pro tyto sekvence dosahovaly nejvyšších hodnot i v případě vlastních vytvořených funkcí, jak již bylo zmíněno výše.

Pro podrobnější analýzu lze využít porovnávání sekvencí pomocí plovoucího okna. Je možné takto odhalit pozitivní selekci, která se v celkové délce řetězce ztratí. Tímto způsobem byl porovnán živočich *Glis Glis* s ostatními organismy. Porovnávaly se geny ND2 a ATP8, u kterých PAML našel významnou hodnotu pozitivní selekce. Při použití plovoucího okna objevil algoritmus pozitivní selekci při porovnání živočich *Glis glis* a *Artibeus jamaicensis* pozitivní selekci v místě okolo pozice 700. Naopak při analýze *Glis glis* a *Cavia porcellus* pozitivní selekce objevena nebyla.

Citovaná literatura

- [1] BALLARD, William O. a David M. RAND. *The Population Biology of Mitochondrial DNA and its Phylogenetic Implications*. Providence: Brown University 2005.
- [2] COOPER, Geoffrey M. *The Cell: A Molecular Approach*. Sunderland, 2000. ISBN 0-87893-106-6.
- [3] CVRČKOVÁ, Fatima. *Úvod do praktické bioinformatiky*. Praha: Academia, 2006. ISBN 80-200-1360-1.
- [4] DEPARTMENT OF STATISTIC, University of Oxford. *Basic models of nucleotide evolution*. [online] Oxford: University of Oxford. 2015.[cit. 2017-05-02]. Dostupné z:
https://www.stats.ox.ac.uk/__data/assets/pdf_file/0003/4296/Basic_Models_of_Nucleotide_Evolution_2.pdf
- [5] FAY, Justin, C., G. J. WYCKOFF a C. WU. *Positive and negative selection on the human genome*. Chicago: University of Chicago 2001. ISBN 1943-2631.
- [6] FELSENSTEIN, Joseph. *Inferring phylogenies*. 2. Sunderland: Sinauer Associates, Inc., 2004. ISBN 978-0878931774.
- [7] Klaban, Vladimír. *Ilustrovaný biologický slovník*. Praha: Galén, 2005. ISBN 80-7262-341-9.
- [8] LOEWE, Laurence. *Negative selection*. [online] Edinburg: University of Edinburgh, , 2008. Dostupné z:
<https://www.nature.com/scitable/topicpage/negative-selection-1136>
- [9] MADĚRÁNKOVÁ, Denisa. *Zarovňávání více sekvencí*. Brno: Vysoké učení technické, 2016. Dostupné z: E-Learning VUT
- [10] NĚMEC, Miroslav a Dagmar MATOULKOVÁ. *Základy obecné mikrobiologie*. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-7923-6.
- [11] REED, Floyd A. *Jukes-Cantor Mutation Model*. [online] University of Hawai'i Reed Lab. 2013.Dostupné z:
<http://hawaiiireedlab.com/wpress/?p=403,%20u%C4%8Debnice%20od%20Ma>d.
- [12] ŘEHOUT, Václav. *Genetická variabilita - mutace*. [online] České Budějovice: Jihočeská univerzita, 2016. [cit. 2017-05-17]. Dostupné z:
http://www.zsf.jcu.cz/cs/katedra/katedra-klinickych-a-preklinickych-oboru/import/ucebni_texty/zaklady-genetiky-a-poradenstvi/12.pdf

- [13] SCOVILLE, Heather. *Synonymous vs. Nonsynonymous mutations*. [online] ©2017.[cit. 2017-04-06]. Dostupné z:
<https://www.thoughtco.com/synonymous-vs-nonsynonymous-mutations-1224600>.
- [14] SCHEFFLER, Immo E. *Mitochondria*. San Diego: University of California, 2008. ISBN 978-0-470-04073-7.
- [15] ŠÍPEK, Antonín. Genetika mitochondrií. *genetika - biologie*. [online] 2014. [cit. 2017-04-06]. Dostupné z: <http://www.genetika-biologie.cz/genetika-mitochondrii>.
- [16] THOMAS, James H. *Detection and Anyalysis of positive selection*. [online] Washington: University of Washington, 2006. [cit. 2017-05-10]. Dostupné z:
<http://depts.washington.edu/jtlab/positiveSelection.html>
- [17] VYHNÁNEK, T. a T. URBAN. *Genové mutace*. [online] Brno: Mendelova univerzita. 2011. [cit. 2017-03-29] Dostupné z:
http://user.mendelu.cz/urban/vsg1/molekul/mol_mutace1.html#Tiché mutace.
- [18] ZHANG, Zhang. *KaKs Calculator: Calculating Ka and Ks Through Model Selection*. [online] Peking: Chinese Academy of Sciences. 2006 [cit. 2017-04-10]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pubmed/17531802>

Seznam zkratek

A	-	Adenin
AK	-	Aminokyseliny
Ala (A)	-	Alinin
Arg (R)	-	Arginin
Asn (N)	-	Asparagin
Asp (D)	-	Kyselina asparagová
ATP	-	Adenosintrifosfát
BEB	-	Bayes empirical Bayes
C	-	Cytosin
CDS	-	Kódující úsek DNA (coding sequence)
Cys (C)	-	Cystein
DNA	-	Deoxyribonukleová kyselina
F81	-	Model Felsenstein
G	-	Guanin
Gln (Q)	-	Glutamin
Glu (E)	-	Kyselina glutamová
Gly (G)	-	Glycin
His (H)	-	Histidin
Ile (I)	-	Isoleucin
JC69	-	Model Jukes-Cantor
K80	-	Dvoupametrický model Kimura
Ka	-	Míra nesynonymních mutací
kb	-	kilobaze
Ks	-	Míra synonymních mutací
Leu (L)	-	Leucin
Lys (K)	-	Lysin
Met (M)	-	Methionin
ML	-	Maximum likelihood
mRNA	-	Mediátorová RNA
msa	-	Vícenásobné zarovnání sekvencí (multi sequence align)
mtDNA	-	Mitochondriální DNA
NC	-	Nukleotid
NCBI	-	National Center for Biotechnology Information
PAML	-	Phylogenetic analysis by maximum likelihood
Phe (F)	-	Fenylalanin
Pro (P)	-	Prolin

RNA	-	Ribonukleová kyselina
rRNA	-	Ribosomální RNA
Ser (S)	-	Serin
T	-	Thymin
Thr (T)	-	Threonin
tRNA	-	Transferová RNA
Trp (W)	-	Trypofan
Tyr (Y)	-	Tytozin
U	-	Uracil
UV	-	Ultrafialové záření
Val (V)	-	Valin

Přílohy

Příloha 1

Tab. 8: Výsledné hodnoty pro *Glis Glis*, gen ATP8

Referenční organismus	Porovnávaný organismus	KaKs	JC69	K80
<i>Glis Glis</i>	<i>Galeopterus variegatus</i>	0,230489	0,415907	0,420886
	<i>Cavia porcellus</i>	0,172119	0,432437	0,441099
	<i>Microtus kikuchii</i>	0,400152	0,424126	0,429590
	<i>Sciurus vulgaris</i>	0,241982	0,415907	0,423269
	<i>Oryctolagus cuniculus</i>	0,270235	0,493340	0,499508
	<i>Lepus europaeus</i>	0,259944	0,540084	0,549891
	<i>Ochotona collaris</i>	0,348673	0,391776	0,393335
	<i>Orycteropus afer</i>	0,313243	0,457935	0,468904
	<i>Macroscelides proboscideus</i>	0,509869	0,449339	0,456485
	<i>Talpa europaea</i>	0,337802	0,383902	0,391466
	<i>Artibeus jamaicensis</i>	0,498267	0,424126	0,440865
	<i>Erinaceus europaeus</i>	0,550893	0,511689	0,551418
	<i>Pteropus dasymallus</i>	0,333546	0,466631	0,476716
	<i>Pteropus scapulatus</i>	0,218727	0,466631	0,473662
	<i>Echinops telfairi</i>	0,315062	0,549793	0,576050
	<i>Tupaia belangeri</i>	0,346466	0,475428	0,488213
	<i>Thryonomys swinderianus</i>	0,232586	0,475428	0,481727
	<i>Isodon macrourus</i>	0,406320	0,589936	0,624790

Příloha 2

Tab. 9: Výsledné hodnoty pro *Glis Glis*, gen ND2

Referenční organismus	Porovnávaný organismus	KaKs	JC69	K80
<i>Glis Glis</i>	<i>Sciurus vulgaris</i>	0,154753	0,465232	0,473938
	<i>Pteropus dasymallus</i>	0,245335	0,484970	0,494394
	<i>Pteropus scapulatus</i>	0,236982	0,501515	0,512478
	<i>Artibeus jamaicensis</i>	0,174576	0,516534	0,534053
	<i>Talpa europaea</i>	0,224000	0,444276	0,456461
	<i>Tupaia belangeri</i>	0,248069	0,486791	0,497427
	<i>Galeopterus variegatus</i>	0,233858	0,486791	0,491301
	<i>Microtus kikuchii</i>	0,292642	0,567534	0,580007
	<i>Erinaceus europaeus</i>	0,205256	0,547507	0,582270
	<i>Isodon macrourus</i>	0,270825	0,573649	0,596575
	<i>Ochotona collaris</i>	0,280562	0,452937	0,465678
	<i>Lepus europaeus</i>	0,179921	0,435715	0,442189
	<i>Oryctolagus cuniculus</i>	0,200131	0,410601	0,418163
	<i>Thryonomys swinderianus</i>	0,280200	0,495960	0,506387
	<i>Cavia porcellus</i>	0,267998	0,444276	0,454697
	<i>Macroscelides proboscideus</i>	0,235290	0,518433	0,530330
	<i>Echinops telfairi</i>	0,237862	0,569567	0,581386
	<i>Orycteropus afer</i>	0,239557	0,499659	0,509715

Příloha 3

Tab. 10: Pozitivní selekce

Organismus, gen	Pozice v sekvenci, AK	Hodnota pozitivní selekce	Hodnota po zprůměrování
<i>Glis Glis, ND2</i>	156, T	1.136**	2.508 +- 0.000
	336, L	1.014**	2.382 +- 0.000
<i>Glis Glis, ATP8</i>	36, P	0.995**	4.478 +- 3.540
	40, T	0.958*	4.325 +- 3.539
	43, D	0.985*	4.459 +- 3.545
	50, H	0.959*	4.327 +- 3.538